



130億パラメータで世界トップクラスの 日本語性能を有する軽量なLLMを開発

日本語LLM*1を開発完了 既にNEC社内で利用開始

NECが開発したファウンデーションモデルの特長



特長①

高い日本語能力



特長②

軽量



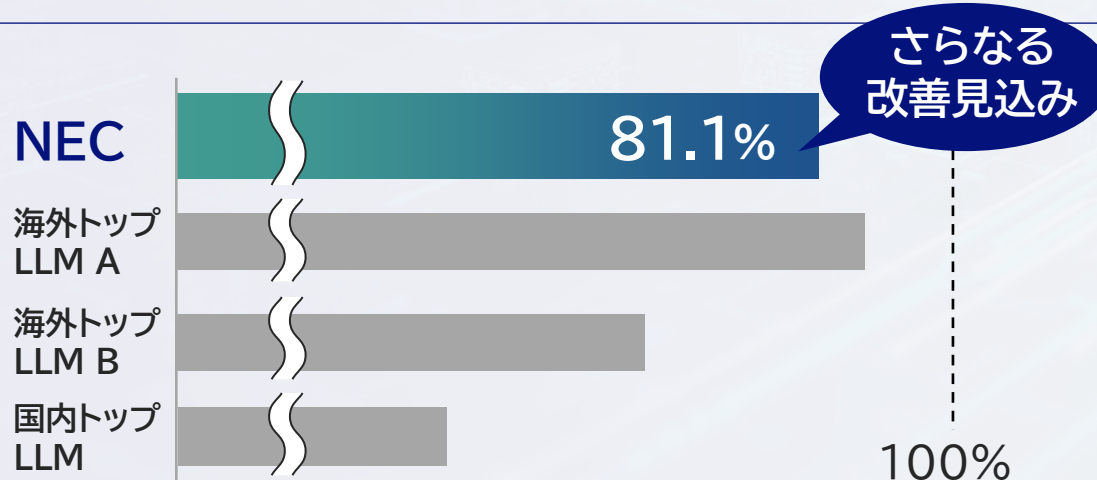
*1 Large Language Model

業務で求められる(1)知識量 (2)文書読解力で
海外トップのLLMに匹敵する性能を達成！

日本語の文書読解力において、トップレベルのLLM 日本語/英語/プログラミング言語が可能

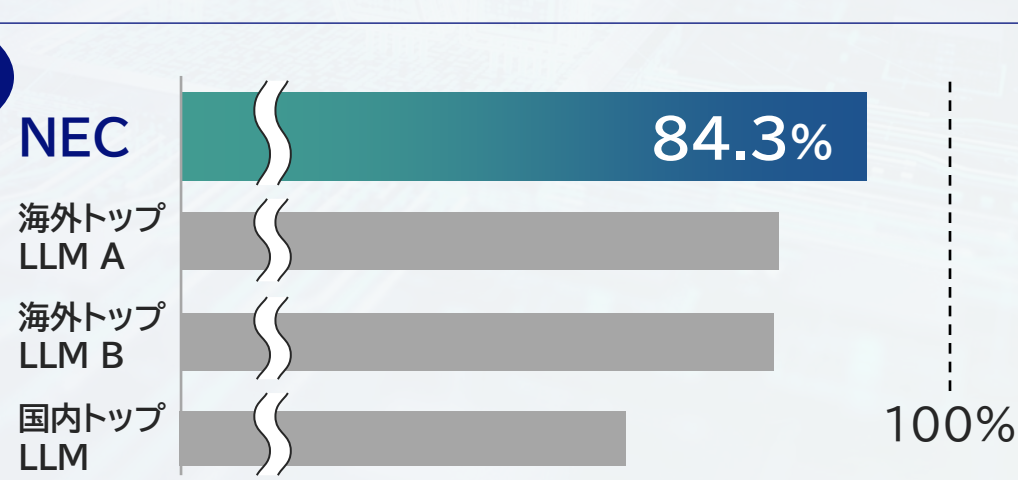
(1)知識量（質問応答）

例) 業務に関する質問に正しく答える力



(2)文書読解力（推論能力）

例) 複雑で長い文章を理解する力



JGLUEベンチマークの JCommonsenseQA, JSQuADスコア
(当社調べ)

特長②
軽量

1/13のモデルサイズ軽量化により コストパフォーマンスを大幅に向上

モデルサイズ
(パラメーターサイズ)



モデルサイズを抑えつつ高い性能を達成

130億



1/13のサイズ
で実現

NEC開発のLLM

軽量化がもたらすお客様のメリット



運用時のサーバコスト・
消費電力を抑制



業務アプリケーションの
レスポンスが高速



オンプレ化が可能
秘匿性の高い業務にも
利用可能



業務にカスタマイズした
顧客専用LLMを
短期間で作成

LLMの性能を最大化するNECのコンピタンス

$$\text{LLMの性能} = \overset{1}{\text{学習データ}} \times \overset{2}{\text{計算機資源}} \times \overset{3}{\text{アルゴリズム}} \times \overset{4}{\text{大規模フィードバック}}$$



NECの
コンピタンス

学習に有効な
データの選別

大量オープンデータから
学習に最適なデータを
選択・加工

国内最大規模の
AI研究用スパコン

大規模な学習を支える
高性能・高信頼な
独自の設計・運用

長年培った
AI技術

グローバルの
多数のAI研究者の
知恵が結集

NEC全社体制で
のフィードバック

幅広い実務利用者の
大規模フィードバックと
それを活用するAI技術

LLM構築を支えるAI研究用スーパーコンピュータ

国内企業では最大規模のAIスパコンが 2023年3月に全面稼働開始

LLM処理向けのAIスパコン構築と運用には、信頼性の確保、性能を引き出すミドルウェア、運用ソフトウェア等が必要であり、NECは研究者を投入し、約2年をかけて構築

AIスパコンの規模

GPU規模 **928基**

演算能力 **580 PFLOPS**



独自AIスパコンにより、実用レベルのファウンデーションモデルを1ヵ月で構築

グローバルでの高い技術競争力 ～AI技術・セキュリティ技術・通信技術～

AI

機械学習
難関国際学会*1 論文採択数
世界企業中10位

映像・画像処理
難関国際学会*2 論文採択数
日本企業中1位

セキュリティ*3 /通信*4

サイバーセキュリティ
山下記念研究賞、CSS2021ほか
論文賞を多数受賞

光通信
難関学会 論文採択
46年連続

特許

企業の特許影響力のグローバル調査
世界の革新的企業トップ100*5
12年連続選出

生体認証＋映像認識＋分析・対応AI
国際特許出願件数*6
世界No.1

機械学習 難関学会採択ランキング（企業）

順位	企業名	文献数
1	Google	1363
2	Microsoft	1342
3	IBM	982
4	DeepMind	450
5	Meta Platforms	428
6	Yahoo	347
7	Alibaba	292
8	Amazon	274
9	Tencent	216
10	NEC	215

2000-2022 当社調べ

*1 NeurIPS, ICML, KDD, ECML-PKDD, ICDM

*2 CVPR, ICCV, ECCV, ACCV, ICPR

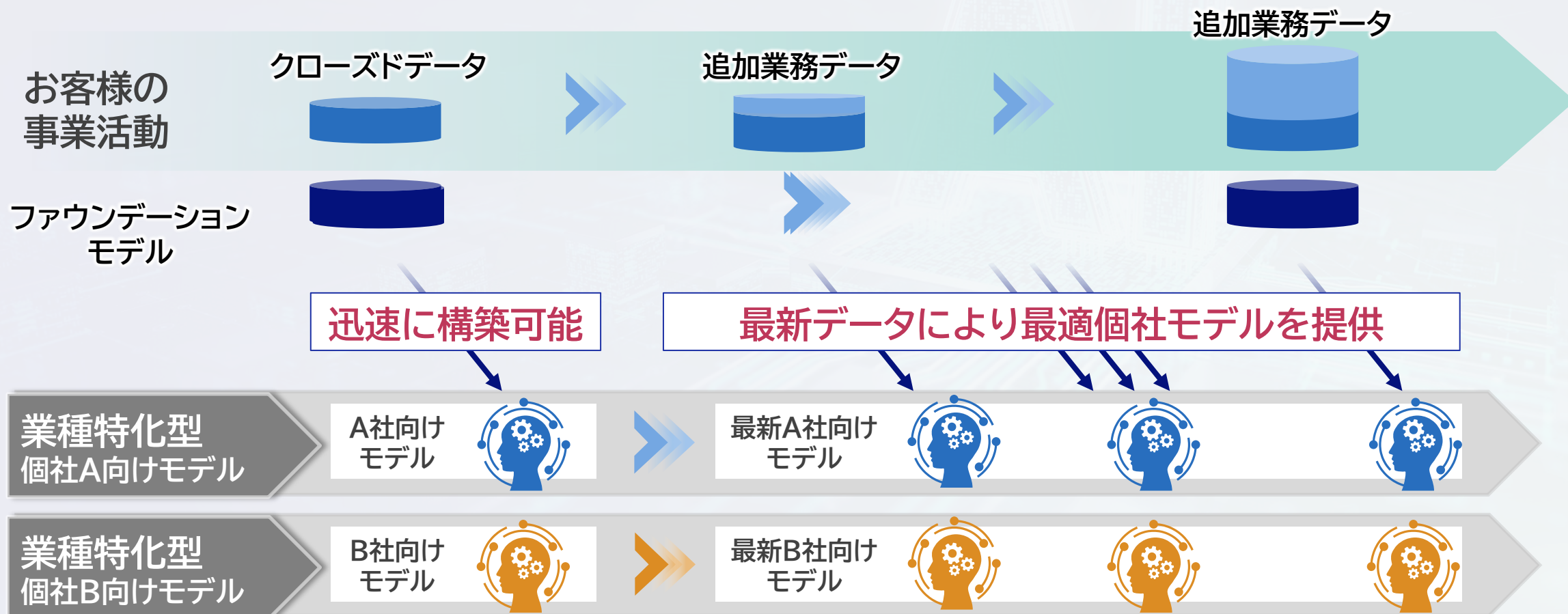
*3 セキュリティ: ACM CCS, Eurocrypt, IEEE S&P等

*4 通信: OFC/ECOC等

*5 TOP100: <https://clarivate.com/top-100-innovators/>

*6 国際特許出願件数: 当社調べ、2022年11月時点での累積PCT出願件数

お客様のクローズドデータを定期的に学習させることにより、最新知見に基づく「個社向けモデル」を継続的にご提供可能

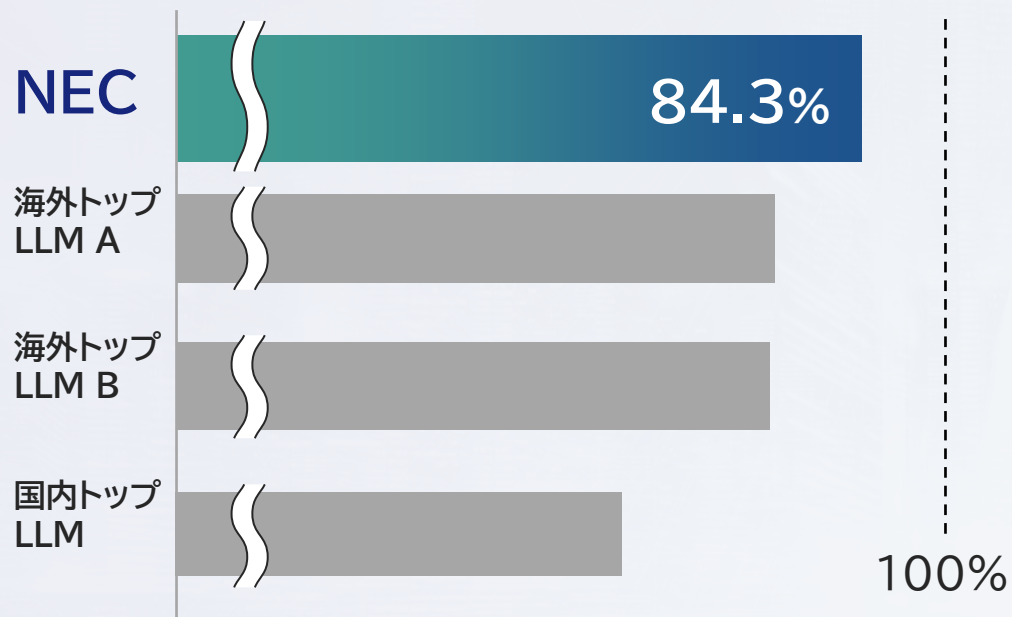


NECのLLMは高い性能とコンパクトさ(軽量)を両立



高い性能(文書読解の能力)

海外トップのLLMを抑えて最高



JGLUEベンチマーク(当社調べ)



コンパクト(軽量)



>1,750億

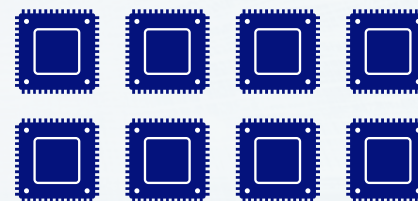


1/13のサイズ
で実現

130億



海外トップLLM



GPU8枚~

NECの
LLM



GPU1枚~

学習データの量/学習時間を増やすことで 高い性能をキープしつつパラメータ数の削減に成功

LLM 性能 = 「学習データ量」 × 「パラメータ数（モデル規模）」（面積）



学習データの量/学習時間を増やすことで 高い性能をキープしつつパラメータ数の削減に成功

LLM 性能 = 「学習データ量」 × 「パラメータ数（モデル規模）」（面積）



学習データの量/学習時間を増やすことで 高い性能をキープしつつパラメータ数の削減に成功

LLM 性能 = 「学習データ量」 × 「パラメータ数（モデル規模）」（面積）



\Orchestrating a brighter world

NEC